LifeCLEF Bird Identification Task 2014

Hervé Goëau, Hervé Glotin, Willem-Pier Vellinga, Robert Planqué, Andreas Rauber, Alexis Joly

LifeCLEF Bird Identification Task 2014

Hervé Goëau[1], Hervé Glotin[2], Willem-Pier Vellinga[3], Robert Planqué[3], Andreas Rauber[4], and Alexis Joly[1,5]

[1] Inria ZENITH team, France, name.surname@inria.fr
[2] Aix Marseille Univ., ENSAM, CNRS LSIS, Univ. Toulon, Institut Univ. de France, glotin@univ-tln.fr
[3] Xeno-canto Foundation, The Netherlands, {wp,bob}@xeno-canto.org
[4] Vienna University of Technology, Austria, rauber@ifs.tuwien.ac.at
[5] LIRMM, Montpellier, France

Abstract. The LifeCLEF bird identification task provides a testbed for a system-oriented evaluation of 501 bird species identification. The main originality of this data is that it was specifically built through a citizen science initiative conducted by Xeno-Canto, an international social network of amateur and expert ornithologists. This makes the task closer to the conditions of a real-world application than previous, similar initiatives. This overview presents the resources and the assessments of the task, summarizes the retrieval approaches employed by the participating groups, and provides an analysis of the main evaluation results. With a total of ten groups from seven countries and with a total of twenty-nine runs submitted, involving distinct and original methods, this first year task confirms the interest of the audio retrieval community for biodiversity and ornithology, and highlights further challenging studies in bird identification.

Keywords: LifeCLEF, bird, song, call, species, retrieval, audio, collection, identification, fine-grained classification, evaluation, benchmark, bioacoustics

1 Introduction

Accurate knowledge of the identity, the geographic distribution and the evolution of bird species is essential for a sustainable development of humanity as well as for biodiversity conservation. Unfortunately, such basic information is often only partially available for professional stakeholders, teachers, scientists and citizens. In fact, it is often incomplete for ecosystems that possess the highest diversity, such as tropical regions. A noticeable cause and consequence of this sparse knowledge is that identifying birds is usually impossible for the general public, and often a difficult task for professionals like park rangers, ecology consultants, and of course, the ornithologists themselves. This "taxonomic gap" [25] was actually identified as one of the main ecological challenges to be solved during United Nations Conference in Rio de Janeiro, Brazil, in 1992.

The use of multimedia identification tools is considered to be one of the most promising solutions to help bridging this taxonomic gap [14], [9], [5], [22], [21]. With the recent advances in digital devices, network bandwidth and information storage capacities, the collection of multimedia data has indeed become an easy task. In parallel, the emergence of "citizen science" and social networking tools has fostered the creation of large and structured communities of nature observers (e.g. eBird[6], Xeno-canto[7], etc.) that have started to produce outstanding collections of multimedia records. Unfortunately, the performance of the state-of-the-art multimedia analysis techniques on such data is still not well understood and it is far from reaching the real world's requirements in terms of identification tools. Most existing studies or available tools typically identify a few tens of species with moderate accuracy whereas they should be scaled-up to take one, two or three orders of magnitude more, in terms of number of species.

The LifeCLEF Bird task proposes to evaluate one of these challenges [12] based on big and real-world data and defined in collaboration with biologists and environmental stakeholders so as to reflect realistic usage scenarios.

Using audio records rather than bird pictures is justified by current practices [5], [22], [21], [4]. Birds are actually not easy to photograph; audio calls and songs have proven to be easier to collect and sufficiently species specific.

Only three notable previous worldwide initiatives on bird species identification based on their songs or calls have taken place, all three in 2013. The first one was the ICML4B bird challenge joint to the International Conference on Machine Learning in Atlanta, June 2013 [2]. It was initiated by the SABIOD MASTODONS CNRS group[8], the University of Toulon and the National Natural History Museum of Paris [10]. It included 35 species, and 76 participants submitted their 400 runs on the Kaggle interface. The second challenge was conducted by F. Brigs at MLSP 2013 workshop, with 15 species, and 79 participants in August 2013. The third challenge, and biggest in 2013, was organised by University of Toulon, SABIOD and Biotope [3], with 80 species from the Provence, France. More than thirty teams participated, reaching 92% of average AUC. Descriptions of the best systems of ICML4B and NIPS4B bird identification challenges are given in the on-line books [2,1] including, in some cases, references to useful scripts.

In collaboration with the organizers of these previous challenges, BirdCLEF 2014 goes one step further by (i) significantly increasing the species number by almost an order of magnitude (ii) working on real-world data collected by hundreds of recordists (iii) moving to a more usage-driven and system-oriented benchmark by allowing the use of meta-data and defining information retrieval oriented metrics. Overall, the task is expected to be much more difficult than previous benchmarks because of the higher confusion risk between the classes, the higher background noise and the higher diversity in the acquisition conditions (devices, recordists uses, contexts diversity, etc.). It will therefore probably produce sub-

[6] http://ebird.org/
[7] http://www.xeno-canto.org/
[8] `http://sabiod.univ-tln.fr`

stantially lower scores and offer a better progression margin towards building real-world generalist identification tools.

2 Dataset

The training and test data of the bird task is composed by audio recordings hosted on Xeno-canto (XC). Xeno-canto is a web-based community of bird sound recordists worldwide with about 1800 active contributors that have already collected more than 175,000 recordings of about 9040 species. 501 species from Brazil are used in the BirdCLEF dataset. They represent the species of that country with the highest number of recordings on XC, totalling 14,027 recordings recorded by hundreds of users. The dataset has between 15 and 91 recordings per species, recorded by between 10 and 42 recordists.

To avoid any bias in the evaluation related to the audio devices used, each audio file has been normalized to a constant bandwidth of 44.1 kHz and coded over 16 bits in .wav mono format (the right channel was selected by default). The conversion from the original Xeno-canto data set was done using ffmpeg, sox and matlab scripts. An optimized 16 Mel Filter Cepstrum Coefficients for bird identification (according to an extended benchmark [7]) have been computed with their first and second temporal derivatives on the whole set. They were used in the best systems run in ICML4B and NIPS4B challenges [2], [1],[3], [10].

Audio records are associated with various meta-data including the species of the most active singing bird, the species of the other birds audible in the background, the type of sound (call, song, alarm, flight, etc.), the date and location of the observations (from which rich statistics on species distribution can be derived), common names and collaborative quality ratings. All of them were produced collaboratively by the Xeno-canto community.

3 Task Description

Participants were asked to determine the species of the most active singing birds in each query file. The background noise can be used as any other meta-data, but it is forbidden to correlate the test set of the challenge with the original annotated Xeno-canto data base (or with any external content as many of them are circulating on the web). More precisely, the whole BirdCLEF dataset has been split in two parts, one for training (and/or indexing) and one for testing. The test set was built by randomly choosing 1/3 of the observations of each species whereas the remaining observations were kept in the reference training set. Recordings of the same species done by the same person the same day are considered as being part of the same observation and cannot be split across the test and training set. The xml files containing the meta-data of the *query* recordings were purged so as to erase the foreground and background species names (the ground truth), the vernacular names (common names of the birds) and the collaborative quality ratings (that would not be available at query stage in a real-world mobile application). Meta-data of the recordings in the training set

are kept unaltered.

The groups participating to the task were asked to produce up to 4 runs containing a ranked list of the most probable species for each record of the test set. Each species had to be associated with a normalized score in the range $[0, 1]$ reflecting the likelihood that this species was singing in the sample. For each submitted run, participants had to say if the run was performed fully automatically or with a human assistance in the processing of the queries, and if they used a method based on only audio analysis or with the use of the metadata. The metric used to compare the runs was the Mean Average Precision averaged across all queries. Since the audio records contain a main species and often some background species belonging to the set of 501 species in the training, we decided to use two metrics, one focusing on all species (MAP1) and a second one focusing only on the main species (MAP2).

4 Participants and methods

87 research groups worldwide registered for the task and downloaded the data (from a total of 127 groups that registered for at least one of the three LifeCLEF tasks). 42 of the 87 registered groups were exclusively registered to the bird task and not to the other LifeCLEF tasks. This shows the high attractiveness of the task in both the multimedia community (presumably interested in several tasks) and in the audio and bioacoustics community (presumably registered only to the bird songs task). Finally, 10 of the 87 registrants, coming from 9 distinct countries, crossed the finish by submitting runs (with a total of 29 runs). These 10 were mainly academics, specialized in bioacoustics, audio processing or multimedia information retrieval. We list them hereafter in alphabetical order and give a brief overview of the techniques they used in their runs. We would like to point out that the LifeCLEF benchmark is a system-oriented evaluation and not a deep or fine evaluation of the underlying algorithms. Readers interested in the scientific and technical details of the implemented methods should refer to the LifeCLEF 2014 working notes or to the research papers of each participant (referenced below):

BiRdSPec, Brazil/Spain, 4 runs: The 4 runs submitted by this group were based on audio features extracted by the Marsyas framework[9] (Time ZeroCrossings features, Spectral Centroid, Flux and Rolloff, and Mel-Frequency Cepstral Coefficients). The runs then differ in two major things: (i) Flat vs. Hierarchical multi-class Support Vector Machine (i.e. using a multi-class Support Vector Machines at each node of the taxonomy as discussed in a research paper of the authors [18]) (ii) classification of full records vs. classification of automatically detected segments (and majority voting on the resulting local predictions). The

[9] http://marsyas.info/

detail of the runs is the following:
BirdSPec Run 1: flat classifier, no segmentation
BirdSPec Run 2: flat classifier, segmentation
BirdSPec Run 3: hierarchical classifier, no segmentation
BirdSPec Run 4: hierarchical classifier, segmentation
Their results (see section 5) show that (i) the segments oriented classification approach brings slight improvements (ii) using the hierarchical classifier does not improve the performances over the flat one (at least using our flat evaluation measure). Note that in every submitted run, only one species was proposed for each query involving lower performances that they should expected with several species propositions.

Golem, Mexico, 3 runs [15]: The audio-only classification method used by this group consists of four stages: (i) pre-processing of the audio signal based on down-sampling and bandpass filtering (between 500hz and 4500hz) (ii) segmentation in syllables (iii) candidate species generation based on HOG features [6] extracted from the syllables and Support Vector Machine (iv) final identification using a Sparse Representation-based classication of HOG features [6] or LBP features [24]. Runs *Golem Run 1* and *Golem Run 2* differ only in the number of candidate species kept at the third stage (100 vs. 50). *Golem Run 3* uses LBP features rather than HOG features for the last step. Best performances were achieved by *Golem Run 1*.

HTL, Singapore, 3 runs [17]: This group experimented several ensembles of classifiers on spectral audio features (filtered MFCC features & spectrum-summarizing features) and metadata features (using 8 fields: Latitude, Longitude, Elevation, Year, Month, Month + Day, Time, Author). The 3 runs mainly differ in the used ensemble of classifiers and the used features:
HLT Run 1: & LDA on audio features locally pooled within 0.5 seconds windows, Random Forest on Metadata (matlab implementation)
HLT Run 2: & LDA, Logistic Regression, SVM, Adaboost and Knn classifier on Metadata and audio features globally pooled with a max pooling strategy, Random Forest on Metadata only (sklearn implementation)
HLT Run 3: & combination of *HLT Run 1* and *HLT Run 2*
Interestingly, in further experiments reported in their working note [17], the authors show that using only the metadata features can perform as well as using only the audio features they experimented.

Inria Zenith, France, 3 runs [11]: This group experimented a fine-grained instance-based classification scheme based on the dense indexing of individual 26-dimensional MFCC features and the pruning of the non-discriminant ones. To make such strategy scalable to the 30M of MFCC features extracted from the tens of thousands audio recordings of the training set, they used high-dimensional hashing techniques coupled with an efficient approximate nearest neighbors search algorithm with controlled quality. Further improvements were

obtained by (i) using a sliding classifier with max pooling (ii) weighting the query features according to their semantic coherence (iii) making use of the metadata to post-filter incoherent species (geo-location, altitude and time-of-day). Runs *INRIA Zenith Run 1* and *INRIA Zenith Run 2* differ in whether the post-filtering based on metadata is used or not.

MNB TSA, Germany, 4 runs [13]: This participant first used the openSMILE audio features extraction tool [8] to extract 57-dimensional low level audio features per frame (35 spectral features, 13 ceptral features, 6 energy features, 3 voicing related features) and then describe an entire audio recording by calculating statistics from the low level features trajectories (as well as their velocity and accelaration trajectories) through 39 functionals including e.g. means, extremes, moments, percentiles and linear as well as quadratic regression. This sums up to 6669-dimensional global features (57 x 3 x 39) per recording that were reduced to 1277-dimensional features through an unsupervised dimension reduction technique. A second type of audio features, namely segment-probabilities, was then extracted. This method consists in using the matching probabilities of segments as features (or more precisely the maxima of the normalized cross-correlation between segments and spectrogrm images using a template matching approach). The details of the different steps including the audio signal preprocessing, the segmentation process and the template matching can be found in [13]. Besides, they also extracted 8 features from the metadata (Year, Month, Time, Latitude, Longitude, Elevation, Locality Index, Author Index). The final classification was done by first selecting the most discriminant features per species (from 100 to 300 features per class) and using the scikit-learn library (ExtraTreesRegressor) for training ensembles of randomized decision trees with probabilistic outputs. Details of the different parameters settings used in each run are detailed in [13]. On average the use of Segment-Probabilities outperforms the other feature sets but for some species the openSMILE and in rare cases even the Metadata feature set was a better choice.

QMUL, UK, 4 runs [19]: This group focused on unsupervised feature learning in order to learn regularities in spectro-temporal content without reference to the training labels and further help the classifier to generalise to further content of the same type. MFCC features and several temporal variants are first extracted from the audio signal after a median-based thresholding pre-processing. Extracted low level features were then reduced through PCA whitening and clustered via spherical k-means (and a two-layer variant of it) to build the vocabulary. During classification, MFCC features are pooled by projecting them on the vocabulary with different temporal pooling strategies. Final supervised classification is achieved thanks to a random forest classifier. This method is the subject of a full-length article which can be read at [20]. Details of the different parameters settings used in each run are detailed in the working note [19].

Randall, France, 1 run: This run *Randall Run 1* is below the ones of the random classifier, which can be explained because of errors in the use of the labels and also by the fact that only one species was proposed for each query, thus this participant did not submit a working note.

SCS, UK, 3 runs [16]: By participating in the LifeCLEF 2014 Bird Task this participant was hoping to demonstrate that spectrogram correlation as implemented in the Ishmael v2.3 library[10] can be very useful for the automatic detection of certain bird calls. Using this method, each test audio record required approximately 12 hours to be processed. The submitted run was consequently restricted to only 14 of the 4339 test audio records, explaining the close to zero evaluation score. This demonstrates the limitation of the approach in the context of large-scale classification.

Utrecht Univ., The Netherlands, 1 run [23] This participant is the only one who experimented with a deep neural network within the task (for the last steps of the method, i.e. feature learning and classification). Their whole framework first includes a decimating and dynamic filtering of the audio signal followed by an energy-based segment detection. Detected segments are then clustered into higher temporal structures through a simple gap-wise merging of smaller sections. MFCC features and several extended variants were then extracted from the consolidated segments before being trained individually by the deep neural network. At query time, an activation-weighted voting strategy was finally used to pool the predictions of the different segments into a final strong classifier.

Yellow Jackets, USA, 1 run As this participant did not submit a working note, we don't have any meaningful information about the submitted run *Yellow Jackets Run 1*. We only know that it achieved very low performances, close to the random classifier. Note that only one species was proposed for each query explaining also these low performances.

Table 1 attempts to summarize the methods used at different stages (feature, classification, subset selection,...) in order to highlight the main choices of participants.

5 Results

Figure 1 and table 1 show the scores obtained by all the runs for the two distinct measured Mean Average Precision (MAP) evaluation measures: MAP1 when considering only the foreground species of each test recording and MAP2 when considering additionally the species listed in the *Background species* field of the metadata. Note that different colors have been used to easily differentiate

[10] http://www.bioacoustics.us/ishmael.html

Table 1. Approaches used by participants. Several sp. in last column indicates if participants gave several ranked species propositions for each query, or if they gave only one species (retrieval vs. pure classification approach).

Team	Preprocessing	Features	Classification	Metadata	Several sp.
BiRdSPec	segmentation (run 4)	Time ZeroCrossings features, Spectral Centroid, Flux and Rolloff, and Mel-Frequency Cepstral Coefficients	"Flat" SVM classifier (run 1 & 2) Taxonomic Hierarchical SVM classifier (run 3 & 4)	Taxonomic hierarchy	×
Golem	down-sampling, bandpass filtering, segmentation in syllables	HOG, LBP	Sparse Representation-based, SVM	×	√
HTL	energy-based segmentation	MFCC, time-averaged spectrograms	Ensemble Classifiers: Logistic Regression, SVM, AdaBoost, Knn, Random Forest, LDA	Latitude, Longitude, Elevation, Year, Month, Month+Day, Time, Author	√
Inria Zenith	noisered specialised filter, (¿ 0:1s) silent passages removed	MFCC	instance-based classification, knnsearch	√ (run 2)	√
MNB TSA	downsampling, noise reduction & segmentation from spectrogram images	Segment-probabilities 57 features (spectral, cepstral, energy, voicing-related features) + velocity + acceleration × 39 statistical functionals	randomized decision trees	Year+month, locality, author	√
QMUL	median-based thresholding	unsupervised feature learning at two time scales	Random forest	×	√
Randall		NA (error in the process)			×
SCS		spectrogram	correlation, 1-nn	×	√
Utrecht Univ.	Segmentation & downsampling	mean MFCC per segment, mean and variance of the MFCCs in a segment, mean, variance and the mean of three sections.	Deep Neural Networks	×	√
Yellow Jackets					×

the methods making use of the metadata from the purely audio-based methods.

Table 2. Raw results of the LifeCLEF 2014 Bird Identification Task

Run name	Type	MAP 1 (with Bg. Sp.)	MAP 2 (without Bg Sp.)
MNB TSA Run 3	AUDIO & METADATA	0,453	0,511
MNB TSA Run 1	AUDIO & METADATA	0,451	0,509
MNB TSA Run 4	AUDIO & METADATA	0,449	0,504
MNB TSA Run 2	AUDIO & METADATA	0,437	0,492
QMUL Run 3	AUDIO	0,355	0,429
QMUL Run 4	AUDIO	0,345	0,414
QMUL Run 2	AUDIO	0,325	0,389
QMUL Run 1	AUDIO	0,308	0,369
INRIA Zenith Run 2	AUDIO & METADATA	0,317	0,365
INRIA Zenith Run 1	AUDIO	0,281	0,328
HLT Run 3	AUDIO & METADATA	0,289	0,272
HLT Run 2	AUDIO & METADATA	0,284	0,267
HLT Run 1	AUDIO & METADATA	0,166	0,159
BirdSPec Run 2	AUDIO	0,119	0,144
Utrecht Univ. Run 1	AUDIO	0,123	0,14
Golem Run 1	AUDIO	0,105	0,129
Golem Run 2	AUDIO	0,104	0,128
BirdSPec Run 1	AUDIO	0,08	0,092
BirdSPec Run 4	AUDIO	0,074	0,089
Golem Run 3	AUDIO	0,074	0,089
BirdSPec Run 3	AUDIO	0,062	0,075
Yellow Jackets Run 1	AUDIO	0,003	0,003
Randall Run 1	AUDIO	0,002	0,002
SCS Run 1	AUDIO	0	0
SCS Run 2	AUDIO	0	0
SCS Run 3	AUDIO	0	0
Perfect Main & Bg. Species	AUDIO	1	0,868
Perfect Main Species	AUDIO	0,784	1
Random Main Species	AUDIO	0,003	0,003

The first main outcome is that the two best performing methods were already among the best performing methods in previous bird identification challenges [2,10,1,3] although LifeCLEF dataset is much bigger and more complex. This clearly demonstrates the generic nature and the stability of the underlying methods. The best performing runs of the MNB TSA group notably confirmed that using matching probabilities of segments as features was once again a good choice. In their working note [13], Lassek et al. actually show that the use of such Segment-Probabilities clearly outperforms the other feature sets they used (0.49 mAP compared to 0.30 for the OpenSmile features [8] and 0.12 for the metadata features). The approach however remains very time consuming as several days

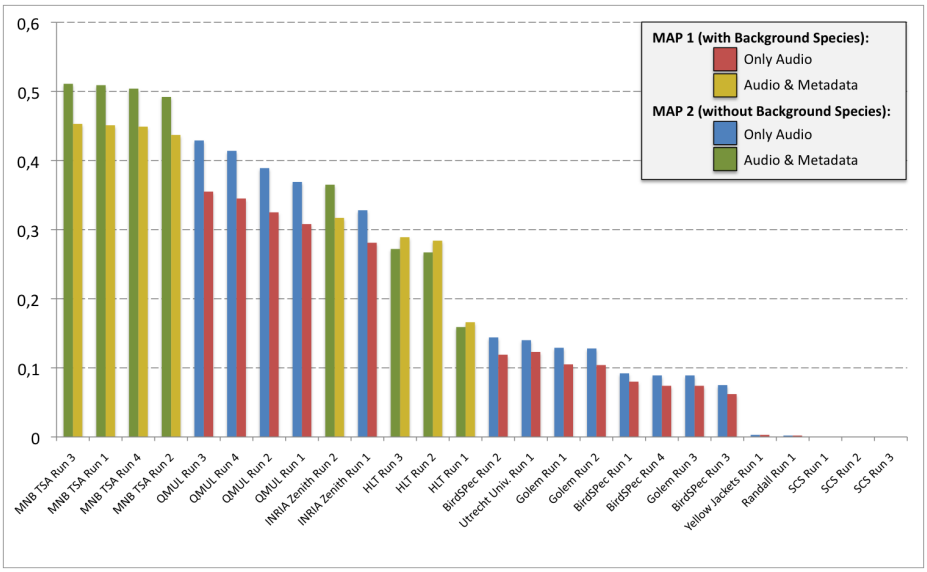

Fig. 1. Official scores of the LifeCLEF Bird Identification Task. MAP1 is the Mean Average Precision averaged across all queries taking int account the Background species (while MAP2 is considering only the foreground species.

on 4 computers were required to process the whole LifeCLEF dataset.
Then, the best performing (purely) audio-based runs of QMUL confirmed that unsupervised feature learning is a simple and effective method to boost classification performance by learning spectro-temporal regularities in the data. They actually show in their working note [19] that their pooling method based on spherical k-means actually produces much more effective features than the raw initial low level features (MFCC based). The principal practical issue with such unsupervised feature learning is that it requires large data volumes to be effective. However, this exhibits a synergy with the large data volumes used within LifeCLEF. This might also explain the rather good performances obtained by the runs of Inria ZENITH group who used hash-based indexing techniques of MFCC features and approximate nearest neigbours classifiers. The underlying hash-based partition and embedding method actually works as an unsupervised feature learning method.
As could be expected, the MAP1 evaluation measure (with the background species) scores are generally lower than the MAP2 scores (without the background species). Only the HTL group did not observe this, and demonstrated the ability of their method to perform a multi-label classification.
A last interesting remark we derived so far from the results comes from the runs submitted by the BirdSPec group. As their two first runs were based on using flat SVM classifiers whereas the 3rd and 4th runs were based on using a hierarchical multi-class SVM classifier it is possible to assess the contribution of

using the taxonomy hierarchy within the classification process. Unfortunately, their results show that this rather tends to slightly degrade the results, at least when using a flat classification evaluation measure as the one we are using. On the other side, we cannot conclude on whether the mistakes done by the flat classifier are further from the correct species compared to the hierarchical one. This would require using a hierarchical evaluation measure (such as the Tree Induced Error) and might be considered in next campaigns.

6 Conclusion

This paper presented the overview and the results of the first LifeCLEF bird identification task. With a number of 87 registrants, it did show a high interest of the multimedia and the bio-accoustic communities in applying their technologies to real-world environmental data such as the ones collected by Xeno-canto. The main outcome of this evaluation is a snapshot of the performances of state-of-the-art techniques that will hopefully serves a guideline for developers interested in building end-user applications. One important conclusion of the campaign is that the two best performing methods were already among the best performing methods in previous bird identification challenges although LifeCLEF dataset is much bigger and more complex. This clearly demonstrates the generic nature of the underlying methods as well as their stability. On the other side, the size of the data was a problem for many registered groups who were not able to produce results within the allocated time and finally abandoned. Even the best performing method of the task (used in the best run) was ran on only 96.8% of the test data and had to be completed by an alternative faster solution for the remaining recordings to be identified. For the next years, we believe is it important to continue working on such large scales and even try to scale up the challenge to thousand species. Maintaining the pressure on the training set size is actually the only way to guaranty that the evaluated technologies could be soon integrated in real-world applications.

References

1. Proc. of Neural Information Processing Scaled for Bioacoustics: from Neurons to Big Data, joint to NIPS (2013), http://sabiod.univ-tln.fr/NIPS4B2013_book.pdf
2. Proc. of the first workshop on Machine Learning for Bioacoustics, joint to ICML (2013), http://sabiod.univ-tln.fr/ICML4B2013_book.pdf
3. Bas, Y., Dufour, O., Glotin, H.: Overview of the nips4b bird classification. In: Proc. of Neural Information Processing Scaled for Bioacoustics: from Neurons to Big Data, joint to NIPS. pp. 12–16 (2013), http://sabiod.univ-tln.fr/NIPS4B2013_book.pdf
4. Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X.Z., Raich, R., Hadley, S.J., Hadley, A.S., Betts, M.G.: Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. The Journal of the Acoustical Society of America 131, 4640 (2012)

5. Cai, J., Ee, D., Pham, B., Roe, P., Zhang, J.: Sensor network for the monitoring of ecosystem: Bird species recognition. In: Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007. 3rd International Conference on. pp. 293–298 (Dec 2007)
6. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. vol. 1, pp. 886–893. IEEE (2005)
7. Dufour, O., Artieres, T., Glotin, H., Giraudet, P.: Clusterized mel filter cepstral coefficients and support vector machines for bird song idenfication. In: Soundscape Semiotics - Localization and Categorization, Glotin (Ed.) (2014)
8. Eyben, F., Wöllmer, M., Schuller, B.: Opensmile: the munich versatile and fast open-source audio feature extractor. In: Proceedings of the international conference on Multimedia. pp. 1459–1462. ACM (2010)
9. Gaston, K.J., O'Neill, M.A.: Automated species identification: why not? Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 359(1444), 655–667 (2004), http://rstb.royalsocietypublishing.org/content/359/1444/655.abstract
10. Glotin, H., Sueur, J.: Overview of the 1st int'l challenge on bird classification. In: Proc. of the first workshop on Machine Learning for Bioacoustics, joint to ICML. pp. 17–21 (2013), http://sabiod.univ-tln.fr/ICML4B2013_book.pdf
11. Joly, A., Champ, J., Buisson, O.: Instance-based bird species identification with undiscriminant features pruning - lifeclef2014. In: Working notes of CLEF 2014 conference (2014)
12. Joly, A., Müller, H., Goëau, H., Glotin, H., Spampinato, C., Rauber, A., Bonnet, P., Vellinga, W.P., Fisher, B.: Lifeclef 2014: multimedia life species identification challenges
13. Lasseck, M.: Large-scale identification of birds in audio recordings. In: Working notes of CLEF 2014 conference (2014)
14. Lee, D.J., Schoenberger, R.B., Shiozawa, D., Xu, X., Zhan, P.: Contour matching for a fish recognition and migration-monitoring system. In: Optics East. pp. 37–48. International Society for Optics and Photonics (2004)
15. Martinez, R., Silvan, L., Villarreal, E.V., Fuentes, G., Meza, I.: Svm candidates and sparse representation for bird identification. In: Working notes of CLEF 2014 conference (2014)
16. Northcott, J.: Overview of the lifeclef 2014 bird task. In: Working notes of CLEF 2014 conference (2014)
17. Ren, L.Y., William Dennis, J., Huy Dat, T.: Bird classification using ensemble classifiers. In: Working notes of CLEF 2014 conference (2014)
18. Silla, C., Freitas, A.: A survey of hierarchical classification across different application domains. Data Mining and Knowledge Discovery 22, 31–72 (2011)
19. Stowell, D., Plumbley, M.D.: Audio-only bird classification using unsupervised feature learning. In: Working notes of CLEF 2014 conference (2014)
20. Stowell, D., Plumbley, M.D.: Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. arXiv preprint arXiv:1405.6524 (2014)
21. Towsey, M., Planitz, B., Nantes, A., Wimmer, J., Roe, P.: A toolbox for animal call recognition. Bioacoustics 21(2), 107–125 (2012)
22. Trifa, V.M., Kirschel, A.N., Taylor, C.E., Vallejo, E.E.: Automated species recognition of antbirds in a mexican rainforest using hidden markov models. The Journal of the Acoustical Society of America 123, 2424 (2008)

23. Vincent Koops, H., van Balen, J., Wiering, F.: A deep neural network approach to the lifeclef 2014 bird task. In: Working notes of CLEF 2014 conference (2014)
24. Wang, L., He, D.C.: Texture classification using texture spectrum. Pattern Recognition 23(8), 905–910 (1990)
25. Wheeler, Q.D., Raven, P.H., Wilson, E.O.: Taxonomy: Impediment or expedient? Science 303(5656), 285 (2004), http://www.sciencemag.org/content/303/5656/285.short